

Achievements and Attitudes in the Learning and Teaching of Mother Tongue and Mathematics in Grades 4 and 9 in the Republic of Azerbaijan

Introduction

The Azerbaijan Educational Reform Project supported by the World Bank is intended to support the efforts of the Azerbaijan Government's Education Reform Programme by improving the effectiveness of teaching and learning in general education. Activities carried out by the consultancy firm Cito will provide the Government with data needed to carry out the 10-year Reform Project in an efficient and relevant way.

The present Azerbaijan National Assessment Study (ANAS) was planned as part of the second stage of Education Reform Project. In the Terms of Reference that cover the present study, mention is made of the establishment of a new system of assessment as one of the main goals in the Educational Reform Project. This is necessary in order to improve teaching and learning, to provide reliable feedback on what students across the country know and can do, and to monitor the national education standards.

National sample-based assessment is to provide stakeholders: parents, educators, policymakers and others with information on the educational system to make sure that children will have the knowledge and skills necessary to be productive members of an improving economy and effective citizens of a developing democracy. For this the nation needs solid, impartial, trusted information about what students know and can do, whether their performance has improved or declined, and the ways they study and learn.

The present ANAS study has surveyed the performance of a representative sample of grade 4 and grade 9 students in general schools in the Republic of Azerbaijan. Their performance was measured in two subjects: Language and Literature (of the Azeri language) and Math in a number of representative content areas (reporting units). Apart from performance data on tests in these two subjects, background data were collected through questionnaires. These were distributed among students and teachers.

The ANAS study has researched the results on the tests, the answers given to the questions in the questionnaire and it has studied the possible relationship between student performance and background variables. The study reports on all findings, discusses some of these in more detail and makes some recommendations in certain areas.

Objectives of the ANAS Study

Assessment of student achievements in various subjects through national assessment has been first carried out by the Consultant in May, 2003. The objective was to pilot national assessment procedures and to develop national assessment

tools. The present Azerbaijan National Assessment Study (ANAS) was planned as part of the second stage of Education Reform Project. In the Terms of Reference (ToR) that cover the present study, mention is made of the establishment of a new system of assessment as one of the main goals in the Educational Reform Project. This is necessary in order to improve teaching and learning, to provide reliable feedback on what students across the country know and can do, and to monitor the national education standards.

The goals of the study as formulated in the ToR are outlined below. The Consultant (Cito) in collaboration with the Student Assessment Unit (SAU) - is to develop and pilot a new national sample-based assessment system by:

- (a) development of guidelines, based on the methodology (concept), for developing instruments for national sample-based assessment and for their implementation in the “general” or “regular” schools; (b) conduct of internal stakeholder review of the guidelines; (c) revision of the guidelines based on the review; (d) submission of the revised guidelines to the MOE for formal review and adoption;
- (a) writing and review of items for the pilot of the national sample-based assessment; (b) development of an item bank at the SAU; (c) conduct of a pilot test in grade 4 and in grade 9; (d) selection of items, in Azeri language (mother tongue) and mathematics, for grades 4 and 9; (e) writing of the student survey items, test booklets and making schemes for the pilot of the national sample-based assessment instruments;
- development of a series of guides for teachers for the pilot of the national sample-based assessment: general guide (national assessment: aims, objectives, technology) and subject/grade specific guides;
- (a) establishment of the design for selecting schools and students for the pilot of the national sample-based assessment; (b) selection of the schools and students;
- (a) establishment of the administration and logistics plan for the pilot of the national sample-based assessment in April-May 2006; (b) preparation of materials needed by school directors, heads of assessment and supervising teachers; (c) printing of the assessment materials and the assessment administration materials in sufficient;
- provision of training for those involved with the pilot of the national sample-based assessment in April-May 2006;
- administration of the pilot of the national sample-based assessment in the Azeri language and mathematics for grades 4 and 9 students; (b) monitoring of the administration; (c) scoring of all assessment instruments; (d) analysis of all results using equipment and software at the SAU; (e) preparation of reports on the results of the pilot of the national sample-based assessment.

Summary of the ANAS Study and Recommendations

Below we will look at the methods used and the results obtained in the Azerbaijan National Assessment Study (ANAS) as presented in the full report of the study. For each area we will give a short summary of the issues reported upon and we will then discuss these. We will end each main section with a list of recommendations.

1 Population, sample and response

1.1 Summary

In the ANAS study two populations were distinguished:

- a population of students: those in grade 4 and those in grade 9;
- a population of teachers: teachers in grade 4 (teaching both Math and Mother Tongue), teachers of Math in grade 9 and teachers of Mother Tongue in grade 9.

Sampling was stratified: first schools were sampled from a given selection of regions, then students were sampled from these schools. Schools were instructed to take part in the study through a letter signed by the Minister of Education that was sent to all schools that had been sampled, a few months prior to both the pre-test and the actual survey. All schools selected took part.

1.2 Discussion and recommendations

The Consultant has discussed with the Ministry of Education whether all regions should take part in ANAS, or that a sample of them would take part. It was decided that for logistic and political reasons a number of regions would not, or in fact, *could* not take part. It is thus possible, or indeed likely, that ANAS does not fully represent the relevant population in the country, even if it does represent a very large part of the country.

It will certainly be possible to generalize about characteristics of and differences between rural and urban areas. However, we cannot be sure if mountainous areas are sufficiently represented by the rural areas that we have sampled. There may very well be factors that are specific for mountainous areas and that may need specific treatment and input as part of the educational reform. In fact, the logistical and financial reasons that mountainous areas were not included in the survey, may influence the organisation of education in these areas generally.

The fact that some areas could not be included in the sample for political reasons is beyond the control of the present project. However, it is a matter of grave concern and it is to be hoped that international organisations will take up this issue and make sure that the educational situation in the areas under occupation will be surveyed as well.

The Terms of Reference for ANAS required us to select *general* schools, which resulted in the exclusion of refugee schools, which are in fact scattered around the country and which would not easily fit in the sampling model that we have used. Yet, given the sizable number of refugee students and refugee schools, it would be of much relevance to carry out a smaller scale survey on refugee students.

The study has profited very much from the cooperation with the Ministry of Education. It is not often that a 100% response rate is reached in this type of surveys.

Summing up, we recommend activities in the following areas:

- Further research into the performance of students and the factors influencing this performance in schools in mountainous areas;
- Contacts with international organisations concerning the state of educational affairs in occupied territories;
- Further research into the performance of students and the factors influencing this performance in refugee schools.

2 Theoretical Framework: justification of choices and conceptual model

2.1 Summary

The following domains in Mother Tongue for grades 4 & 9 were selected to be reported on in reporting units:

- Grammar (morphology, syntax)
- Reading (everyday texts and fictional/literary texts).

For Math the following domains were selected to be reported on in reporting units:

Grade 4:

- Quantities
- Addition and Subtraction
- Multiplication and Division

Grade 9:

- Algebra
- Geometry

For each of these domains, specific subjects were chosen. When ANAS was carried out, the old teaching plan was still in force and a new curriculum was being developed. The domains and the items constructed for them are therefore based on the old teaching plan.

As to the conceptual model, the main objective of ANAS was to measure student achievements in two different subjects Math and Mother Tongue and to identify how various factors affect achievements. Factors that may affect achievement are given below.

Student factors

- basic student characteristics
- student's educational career
- home context and family background
- out of school activities
- attitude towards Math or Mother Tongue

Class/teacher factors

- basic class characteristics
- instruction and learning
- assessment and homework policies and activities
- classroom environment, structure and climate
- teacher characteristics, education and preparation

School factors

- basic school characteristics / demographics
- school resources
- school climate
- school policies and practices

2.2 Discussion and recommendations

When ANAS was carried out, the old teaching plan was still in force and a new curriculum was being developed. The fact that the survey does not contain items that specifically tap (sub)skills as defined in the new curriculum, may create some problems in later years. The present study is very much like a baseline study: it tells us what the performance level of students is and what factors affect these, with the old teaching plan still in force. However, if we are to compare the 2006 findings with findings in a repeat study some 5 years later, we need to be able to compare students' achievements in comparable skills that have been taught to them. We have tried to overcome this issue by carefully studying the plans and the conditions for introducing the new curriculum and identifying those domains that would probably be included in the new curriculum and were present in the old teaching plan. There were some problems relating teaching plan and curriculum: the

relationship between the standards in the new curriculum which is being developed and the previous national standards is not always clearly visible. Although the standards in the newly developed curriculum seem to be more concrete, it has not been indicated to which general objective each of them serves. In the standards the level of mastery must be related to the extent to which students are able to apply the knowledge, skills and abilities gained in that subject to another subjects (rather than to the percentage of students who can do this). I

Once the new curriculum has been in force for some years, a repeat study will need to be prepared to measure progress, i.e. to see whether the reform of the educational system has had the outcomes that it was supposed to have. In this repeat study we can include the same questions as we did in the 2006 study. However, there are also ways to include new tasks and questions that are more specific to the new curriculum, while retaining enough tasks and questions to relate to the performance under the old teaching programme. For such an operation a different test design is necessary.

As to the conceptual framework, we have identified those factors that tend to affect student performance. We have based ourselves on models in use in international assessment such as PISA and TIMSS. We have also taken into account the framework that was used in the baseline study carried out in 2004 by the same Consultant. For logistic reasons (students only have a limited period of time to answer questions) we have had to exclude other factors that may influence student behaviour. One issue is that Azerbaijan is going through a very rapid economic development process: the economy is booming and more money will be available to on the hand modernize facilities at school and on the other hand improve living conditions in the homes. Although the basic rubrics in the conceptual framework will probably remain the same, it would be useful to examine if other factors may have to be added or may have to replace present ones.

Summing up, we recommend activities in the following areas:

- A detailed study or inventory of the differences between the old teaching programme and the new curriculum would greatly help the definition of domains and tasks for the preparation of the repeat study to measure progress.
- On the basis of the study mentioned above, new tasks and questions may need to be developed which tap the (sub)skills to be developed under the new curriculum in a more specific way.
- A limited revision of the conceptual framework may be necessary when Azerbaijan has been through economic changes.

3 Development of instruments

3.1 Summary

Construction of items

Items were constructed in line with objectives developed by the working groups. Three item types were distinguished: multiple choice, short answer and open-ended questions. A total of 70 minutes was assigned for responding to items per test. It was checked if the (number of) items selected would not put too much strain on the students.

After the pre-test and analysis, for grade 4 Math 50 items were selected for inclusion in the final test, for grade 9 the number of items was 44. For Mother Tongue, the same procedure was followed, resulting in 74 items for grade 4 and 75 items for grade 9.

Together with the construction of the items, marking schemes were developed. These were checked by Cito consultants for possible inconsistencies and for efficiency at data entering. The marking schemes were trialled during the pre-test round.

Questionnaires

Questionnaires were constructed to identify the extent, to which the factors identified in the conceptual model affect student achievements. Questionnaires were developed for students and for teachers, for Mother Tongue and for Math, in grades 4 and 9. In grade 4 teachers teach both Math and Mother Tongue, so only one version of the questionnaire was produced. For grade 9 teacher questionnaires there were two versions. Questionnaires have been pre-tested together with the items. Student questionnaires have been pre-tested together with the items. 10 minutes were allocated for completing questionnaires.

Instructions for test administration

Given the sensitive nature of ANAS, it was decided to assign a *test administrator* for each school. Supervisors were also selected among participants of earlier seminars, school teachers and staff of the Institute for Educational Problems, the Azerbaijan Teachers Institute and the Institute for the Improvement and Re-education of Pedagogical Personnel. For all these functionaries, including school principals, separate guidelines were developed.

Prior to the administration of ANAS, the ANAS coordinator (a national consultant appointed by the Consultant) visited the relevant regional centres to instruct supervisors and hand out guidelines for making them closely acquainted with procedures of ANAS. Workshops were held for school principals, test administrators and supervisors.

3.2 Discussion and recommendations

Cito consultants and national consultants have been fruitfully cooperating and have been able to produce instruments that comply with international standards, as was shown in both the pre-test and in the final test.

It was a great help that national consultants had gained much expertise in similar activities in 2002 and 2004 in projects led by the Consultant. Test construction was made difficult by the fact that the new curriculum was not available yet, the construction of items itself did not pose any serious problems.

Special mention must be made of the ANAS coordinator in the project, who has successfully coordinated all the activities in this area and has personally given all the trainings that were mentioned above.

The Consultants have some concern that none of the experienced consultants mentioned above is a staff member in the Student Assessment Unit in the Ministry and that none of the regular staff of that Unit has been involved in the activities described above (because of other duties). Although every activity has been meticulously documented we feel that there is a grave danger of expertise being lost if in five years time the study has to be repeated.

Summing up, we recommend activities in the following areas:

- Expanding the Student Assessment Unit to include more staff to help prepare and carry out national assessments.
- Putting measures in place to keep expertise in this area available.

4 Administrative design

4.1 Summary

In order to successfully carry out the survey a number of activities have been undertaken:

- Coding of test booklets: test booklets were coded (12 digits indicating such variables as region, district, school, grade and student) and booklets were packed in the central office.
- Mother Tongue tests were in one booklet consisting of two parts (Reading and Grammar). Total testing time was 90 minutes with a five minute break between the two parts.
- Math tests were also in two parts: for grade 4 the three domains (Quantities; Addition and Subtraction; Multiplication and Division) were divided across the two parts, for grade 9 one part contained Algebra and the other Geometry.
- Various forms were produced to provide the necessary documentation during ANAS. Forms for recording the delivery and the returning of questionnaires, for reporting on test administration at schools, on the testing process, and for data about education of parents and social status of families.
- Packaging of test booklets was necessary to transfer them to schools. Test booklets and questionnaires were placed in envelopes in the required numbers with some spare copies in each envelope.
- Administration of ANAS was implemented from 21 April – 13 May 2006 in the various regions selected, following the decree of the Minister of Education. The survey took two days to administer at each school. The Math test was administered on the first day and the Mother Tongue test on the second day. The allocated time period was sufficient for most of students. They indicated that they needed an average of 30-40 minutes to respond to the items.

4.2 Discussion and recommendations

Thanks to the experience of the national consultants hired by Cito, the whole process of the administration of the survey has been successful. Given the limited equipment and facilities available for this project and the sometimes difficult circumstances of reaching far-away schools, the administration can be called very successful, with no irregularities having been reported.

Again, we like to stress that expertise in this area must be cherished.

Summing up, we repeat our earlier recommendations on activities in the following areas:

- Expanding the Student Assessment Unit to include more staff to help prepare and carry out national assessments.
- Putting measures in place to keep expertise in this area available.

5 Rating, data entry and data cleaning

5.1 Summary

Various activities have taken place in this area:

- Guidelines for rating have been developed according to the subject and the type of items. Responses to multiple choice items were to be entered into the computer using dedicated software. Identification and scoring of correct answers to short answer and open questions were also presented in the guidelines.

- Rating of test items was carried out by national consultants. In cases that students produced unexpected answers, the working group discussed the issue and took an appropriate decision.
- Data entry for questionnaires, test items and student background (parental education, social status) were entered into the computer. Specific forms were developed for this in a Microsoft Excel format. Eight forms were developed for data entry.
Persons hired for entering data were checked regularly. The verification/checking of the data entry process was carried out with the help of Overton 32 computer software that was made available by the Cito international consultants.
- Data cleaning procedures were used to guarantee the quality of the data set. Data cleaning involved two steps: *a validation procedure* where for each entry a list of valid codes was defined and *a check on inconsistent data*: where all data were checked for inconsistency within each questionnaire. All original data files were saved in order to keep track of all changes that were made during data cleaning.

5.2 Discussion and recommendations

Data cleaning is a complicated process that is an essential part of research like the present survey. Given the fact that there was a variety of data in different formats (scores on tests containing various item types) and questions in questionnaires in various formats, quality control is essential. There has been frequent communication between the Cito consultant and the ANAS coordinator cum psychometrician on inconsistent data. This communication was made difficult by the fact the two parties did not share a common language.

Given the fact that the literature in this area is in English mainly, it would be advisable for this national consultant to be trained in English more extensively. Another training that this expert would much profit from, is in various methods of data cleaning, quality control and quality assurance. It is the Consultant's impression that the present expert is one of the very few experts in Azerbaijan in this field. Yet we are not sure that this consultant would be able to operate independently in a next round of national assessment.

Summing up, we recommend activities in the following areas:

- The local expert to be trained extensively in the English language.
- The local expert to be trained in aspects of data cleaning, quality control and quality assurance.

6 Statistical Analysis

6.1 Summary

Data were analysed with the help of software:

- *TIA-plus*: classical test and item analysis was carried out with the help of this software, developed at Cito and made available to the project.
- *SPSS*: data precision, statistical analysis of questions in questionnaires, correlation analysis on variables, performance in comparison with variables, regression analysis.

6.2 Discussion and recommendations

It was found that the ANAS coordinator was well-versed in classical test analyses and that TIA-plus functioned well in the hands of the coordinator. In the case of SPSS it was found that the coordinator could not function independently of the Cito consultant and that further training in this area is needed. This was originally envisioned in the Terms of Reference to this project where the local expert would have been offered a training course abroad for a considerable period of time. For internal reasons the Ministry decided to remove this training from the project plan, which left the project rather handicapped at this stage of the study. As with other activities the local expert was not able to function independently. This may cause serious problems at a later stage when the survey will have to be repeated to measure progress.

There is another reason for further training of the local expert in this area. This has resulted in questions and tasks in the test that tap the present teaching programme sufficiently but that may not be pertinent enough to the newly developed curriculum. In the repeat study in circa five years from now, it would be highly advisable to include more questions relating to the new curriculum then in force. However, this would mean in the present test design that either the test would have to be lengthened or tasks would have to be dropped to be replaced by new ones. Both solutions would weaken the strength of the next survey and its ability to measure progress.

There is a solution to this problem that is generally applied in international surveys: the application of incomplete designs where students would not each be administered the same test, but instead a number of overlapping subtests would be administered to students. Each subtest would contain a large part of the original questions plus some new questions that tap skills more related to the new curriculum. To analyse data collected in such an incomplete design, a type of analysis is needed that is different from the classical test and item analysis that has been used until now. This alternative would be analysis based on Item Response Theory (IRT). The local expert is aware of the importance of IRT and indeed the Ministry in their original Terms of Reference would have had a local expert trained in this type of analysis. Thus, for a number of reasons further training of a local expert in statistics is needed.

Summing up, we recommend activities in the following areas:

- Investigating possibilities to apply incomplete designs in surveys to come.
- The local expert to be trained in IRT based analysis.

7 Standard Setting

7.1 Summary

Azeri national consultants in the ANAS study have been given an introduction by the Consultant to the principles of standard setting. A standard setting procedure for sample-based assessment is the application of a rational system of rules and procedures resulting in a test score that differentiates between levels of performance. In the case of sample-based assessment experts will need to determine which goals as operationalised in the items within a reporting unit need to be mastered by the whole population (100%) and by a smaller group (30%) of talented students, who may wish to pursue their studies in specialized schools after grade 9. In the case of sample-based assessment we have had the percentages of correct answers available on each item in the test. This has enabled experts to set more realistic standards than in the case of curriculum based standards setting where such information is not (yet) available.

7.2 Discussion and Recommendations

We have described how the Azeri national consultants in the ANAS research have taken part in a *trial* standard setting procedure. The issue is here that the ANAS consultants cannot be seen to represent the Ministry's views on what needs to be achieved by which percentage of students. During the ANAS study the Ministry had another more urgent issue to address, namely the development of *new* curriculum content standards themselves. One these have been finalized it seems proper to do a standard setting on the basis of the results of the ANAS research.

Summing up, we recommend activities in the following areas:

- Carrying out a standard setting procedure with specialists who have been engaged in developing the new curriculum and assessment standards on the basis of the data of student performance as collected in the ANAS study.

8 Performance in Reporting Units: Mother Tongue

8.1. Summary

The following is a summary of the results on the Mother Tongue test at grades 4 & 9.

Grade 4

The grade 4 test booklets consisted of two parts: Grammar (50 items) and Reading (24 items). In the Grammar part the domains of morphology and syntax were distinguished. In the Reading part there were also two domains: the reading of everyday texts and the reading of fictional/literary texts.

In the following tables we show a summary of the performance on the test and parts for the whole population, the regions, and urban and rural areas.

Part of test	Score	P-value
Complete test	45	60
Grammar	30	61
Morphology	21	64
Syntax	10	58
Reading (total)	14	58
Reading (everyday texts)	4	53
Reading (fictional texts)	10	61

Table 1 *Scores and P-values of the total population on the complete grade 4 test of Mother Tongue*

Region	Score	P-value
Absheron	44	60
Baku	52	70
Gence-Qazax	42	57
Aran	41	55
Sheki-Zaqalatala	47	63
Dagliq Shirvan	36	48
Lenkeran-Astara	39	53
Quba-Xachmaz	41	56

TOTALS	45	60
--------	----	----

Table 2 Mean scores and p-values on the complete grade 4 test of Mother Tongue per region

Geographical area	Score	P-value
Urban	48	65
Rural	38	52

Table 3 Mean Scores on the grade 9 test of Mother Tongue urban vs. regional areas

Grade 9

The grade 9 test booklets consisted of two parts: Grammar (50 items) and Reading (25 items). In the Grammar part the domains of morphology and syntax were distinguished. In the Reading part there were also two domains: the reading of everyday texts and the reading of fictional/literary texts.

Part of test	Score	P-value
Complete test	30	40
Grammar	18	36
Sentence Analysis	14	36
Indirect/direct speech	5	37
Reading (total)	12	48
Reading (everyday texts)	5	46
Reading (fictional texts)	7	51

Table 4 Mean scores and P-values of the total population on the complete grade 9 test of Mother Tongue

Region	Score	P-value
Total Population	30	40
Absheron	29	38
Baku	37	49
Gence-Qazax	27	36
Aran	27	33
Sheki-Zaqalatala	23	37
Dagliq Shirvan	18	38
Lenkeran-Astara	28	38
Quba-Xachmaz	30	40

Table 5 Mean Scores and P-values on the complete grade 9 test of Mother Tongue per region

Geographical area	Score	P-value
Urban	32	45
Rural	26	32

Table 6 Mean Scores and P-values on the grade 9 test of Mother Tongue urban vs. regional areas

8.2 Discussion and Recommendations

The results of the two populations show that there is a difference in performance between rural and urban areas and that indeed the Baku population scores highest in both grades. This confirms earlier findings from the 2002 study carried out by the Consultant. In the baseline study carried out by the Consultant in 2004, it was found that facilities in the urban areas were better developed than in the rural areas. The

availability of better equipment, materials and better qualified staff may be part of the explanation for the higher performance in urban areas. We have collected some background variables that link up with the 2004 study. It would be well worth carrying out further research into the extent to which the availability of facilities indeed is the (main) explanation for the observed difference in performance.

It is worth noting that students in both grades are better at reading fictional texts than so-called every-day texts. The latter are the type of functional texts that any reader of the ages concerned will come across in newspapers, magazines, notices etc.

Traditionally, fictional and literary tests play a major part in the language and literature classes in Azerbaijan and this may indeed be the reason that students are better in this area of reading. However, in international surveys such as PISA and PIRLS functional tests of the type mentioned above are very much part of the construct of reading literacy as tested in these surveys. It would be interesting to see if the results on the PISA 2006 survey as carried out in Azerbaijan will confirm the findings of the present survey.

Summing up, we recommend activities in the following areas:

- Carrying out further research into the extent to which the availability of facilities is an explanation for the observed difference in performance between rural and urban areas.
- Carrying out comparative studies between the results on the PISA 2006 survey and the present survey (for grade 9).

9 Performance in Reporting Units: Math

9.1. Summary

The following is a summary of the results on the Math test at grades 4 & 9.

Grade 4

The grade 4 test booklets contained a total of 50 tasks in three domains: Quantities, Addition and Subtraction, and Multiplication and Division,

In the following tables we show a summary of the performance on the test and parts for the whole population, the regions, and urban and rural areas.

Part of test	Score	P-value
Complete test	52	56
Quantities	6	58
Addition and Subtraction	9	52
Multiplication and Division	36	57

Table 7 Mean scores and P-values of the total population on the complete grade 4 test of Math

Region	Score	P-value
Absheron	53	58
Baku	61	66
Gence-Qazax	49	54
Aran	47	51
Sheki-Zaqalatala	51	56
Dagliq Shirvan	40	44

Lenkeran-Astara	40	45
Quba-Xachmaz	47	51

Table 8 Mean Scores and P-values on the complete grade 4 test of Math per region

Geographical area	Score	P-value
Urban	56	61
Rural	43	47

Table 9 Mean Scores and P-values on the grade 4 test of Math urban vs. regional areas

Grade 9

The grade 9 test booklets contained a total of 44 tasks in two domains: Algebra and Geometry.

In the following tables we show a summary of the performance on the test and parts for the whole population, the regions, and urban and rural areas.

Part of test	Score	P-value
Complete test	22	30
geometry	5	28
Algebra	15	31

Table 10 Mean Scores and P-values of the total population on the complete grade 9 test of Math

Region	Score	P-value
Absheron	26	35
Baku	27	37
Gence-Qazax	19	25
Aran	18	24
Sheki-Zaqalatala	16	21
Dagliq Shirvan	30	40
Lenkeran-Astara	18	24
Quba-Xachmaz	23	30

Table 11 Mean Scores and P-values on the complete grade 9 test of Math per region

Geographical area	Score	P-value
Urban	25	33
Rural	18	24

Table 12 Mean Scores and P-values on the grade 9 test of Math urban vs. regional areas

9.2 Discussion and Recommendations

As with Mother Tongue, the results of the two populations show that there is a difference in performance between rural and urban areas with the latter scoring highest in both grades. Again this confirms findings in earlier studies carried out by the Consultant in 2002 and 2004.

However, looking at the results per region it is remarkable that the typically rural area of Dagliq Shirvan has the highest score in the grade 9 test. This is all the more remarkable when we see that that region is among the lowest performers both in grade 4 Math and in Mother Tongue grades 4 & 9.

The Consultant and the local ANAS coordinator have carefully checked all the data, but for the time being we cannot find any irregularity in the data collection, data processing and data analysis. Yet we feel that we must postpone our judgement of this result.

Like for Mother Tongue, the availability of better equipment, materials and better qualified staff may be part of the explanation for the higher performance in urban areas. We have collected some background variables that link up with the 2004 study. It would be well worth carrying out further research into the extent to which the availability of facilities offers an explanation for the observed difference in performance.

Summing up, we recommend activities in the following areas:

- Carrying out further research into the extent to which the availability of facilities is an explanation for the observed difference in performance between rural and urban areas.
- Carrying out further research into the unexpectedly high performance of students in the Dagliq Shirvan in the test for grade 9.

10 Questionnaires

10.1 Summary

The following is a summary of the data collected on the basis of the questionnaires that were distributed to all sampled students in grades 4 & 9, their general teachers in grade 4 and their Math and Mother Tongue teachers in grade 9.

Student background variables

Data were collected on geographical background, age and gender of students. In grade 4 the largest section of students was from an urban background (67%) and was 10 years old (52%). Girls and boys were equally represented. For grade 9 we found the following. Urban background: 64%, the largest age-section was 15 years old: 52% and again an equal representation of boys (49%) and girls.

Home background and family background of students

Most students are brought up by the two parents (over 80% for both grades) and Azeri is the first language spoken in over 95% of homes. For both grade-groups the great majority of parents has had secondary education (over 70%).

We have seen that over 60% of students in grade 9 have a calculator at their disposal, which may be necessary to their homework for Math and possibly other subjects. In grade 4, where this need is a little less urgent, only just over 40% of students have a calculator available.

Computers are not much available: around 20% of homes have a computer.

Private text books, that is: school books to learn and work from, are available in ca 80% of homes. This may be an important condition for success in education: students are free to take their books home and study from them when there is a need for this. A quiet place to study is another such favourable condition for success in school. Over 80% of homes have such a facility.

In the perception of over 35% of students in grade 4 and over 15% of those in grade 9, there are no books in the home. We will comment on this issue in the section on discussion and recommendations below.

Out of school activities

A small percentage (8%) of students in grade 4 carries out activities for which money is received. 85% of students have to do household chores. Apart from this, students carry out a large variety of leisure activities. Students in grade 9 also have to assist in household chores. Some 26% of students are engaged in a job outside the house for which money is received. This job may take from less than one hour to more than five hours per day. Leisure activities vary widely. Given the scarcity of computers in

the home (compared to many countries elsewhere in Europe), it is no surprise that 57% of students never plays a computer game.

Attitude to Mother Tongue and Math

Students in grade 4 are more positive to both subjects than students in grade 9. In this grade 18% students do not like Math and 11% do not like Math *at all*. The general attitude of grade 9 students to both subjects remains positive. A total of over 80% agree or agree very much that Math is necessary for further education in university and in later life. For Mother Tongue this percentage is even higher: at around 90%.

Instruction and learning

According to the students, most teaching methods, as identified in the questionnaires, were practised by the teachers *regularly* to *often*. Working in small groups occurred least frequently: according to 16% of grade 4 students this never happened. According to 33% of students in grade 9 this never happens in Mother Tongue lessons, and according to 29% of students this never happens in grade 9 Math lessons.

According to the teachers, a great variety of teaching methods is being practised. There is a difference between teachers' and student's views as to the frequency of working in small groups, which the teachers in grade 4 never practise only according to 3% of teachers (cf 16% of students). 3% of teachers of Math and 4% of Mother Tongue teachers reply that they never make students work in small groups, as compared to the opinion of 29% and 33% of students that indicate that this never happens.

Assessment and homework policies and activities

In both grades a large variety of assessment methods is used, of which setting oral questions is the most popular: in grade it is practised in every lesson by over 80% of teachers. The same is true of grade 9 teachers.

The student answers indicate that teachers frequently assess their pupils, with the checking of homework being the most frequent method: 78% of grade 4 students say that this always happens. In grade 9 for Mother Tongue this is true according to 78% of students and for Math according to 69% of students.

A great variety of learning activities as identified in the questionnaires is set by teachers in both grades. Comparatively less time seems to be spent on implementing projects and in practical exercises: 23% of teachers say they do this in every lesson. In grade 9 this is done by 40% of Mother Tongue teachers and by 39% of Math teachers in every lesson.

In grade 4 around 70% of teachers set homework every day for both Mother Tongue and Math. In grade 9, 80% of Math teachers set homework every day and 82% of Mother Tongue teachers do this. Around 60% of teachers check the homework of the students up to 5 minutes per lesson, with a smaller percentage of between 20 and 30 of teachers checking homework more than 5 minutes per lesson.

Classroom environment

The appreciation by grade 4 students of the attitude of teachers to students is positive: over 90% find that the teacher cares for the students. Their appreciation of fellow-students is similar. Some 70% of grade 9 students experience that teachers encourage students to solve problems on their own. As to order in class, around 20% of students find that teachers need a long time to make students silent.

A similar positive attitude is found among teachers: over 90% of teachers in grade 4 and over 80% of teachers in grade 9 find that that students often or always carry out exactly what they instruct and that students work systematically and quietly. Over 95 % of teachers think that students often or always tend to approach their assignments with seriousness.

Teacher characteristics, education and preparation

Nearly half the number of teachers in grade 4 is over 46 years old, with the largest age group being those of 46-55 years old: 30%. The smallest age group (4%) is that of the 22-25 year-olds. Most teachers in grade 9 are between 36 and 55 years old. In age group 36-45 there are 24% of teachers and in age group 46-55 there are 43% of teachers.

In both grades female teachers are in the majority, in grade 4 by 94%, in grade 9 by 72%.

All teachers in grade 4 and over 99 % of teachers in grade 9 have had a pedagogical training. In grade 62% of teachers had a higher pedagogical training at a university or equivalent and 39% of teachers had a training at a specialized secondary college. In grade 9 over 99% all teachers have had a training at a higher pedagogical college. In grade 4, bachelor degrees are held by 7% and master degrees by 4% of teachers. In grade 9, bachelor degrees are held by 17% and master degrees by 19% of teachers

Both in grade 4 and in grade 9 teachers have a long work experience, with 37% of teachers in grade 4 and 46% of teachers in grade 9 having worked in education for more than 25 years. In grade 4 97% of teachers have worked in education for longer than 3 years. In grade 9 nearly 90% of teachers have worked in education for longer than 3 years.

School resources

A great variety of resources is available for students during lessons, of which private textbooks are available for over 80% of students in both grades. Computers and internet are less available: in grade 4 between 15% and 18% for Mother Tongue and math lessons, and in grade 9 for 13 % of students in both subjects.

School climate

76% of teachers in grade 4 and 71% of teachers in grade 9 think that their colleagues teach the full time that they are supposed to teach. When they compare results in other classes with the results in their own classes 39% of grade 4 teachers and 38% of grade 9 teachers think that their own results are never lower than in other classes. The majority of teachers find that students are disciplined, with a total of 23% of teachers in grade 4 thinking students that discipline is lacking to some degree. In grade 9 over 50% of teachers are satisfied that students are disciplined, with varying percentages of teachers indicating that this is not always so for some teachers. Homogeneity in classes is to be found in both grades and both subjects: 74% of grade 4 teachers think this is applicable to a varying degree. In grade 9, 78% of Math teachers and 81% of Mother Tongue teachers think this.

10.2 Discussion and Recommendations

In this section we will mainly look at the data mentioned above without reference to scores on the test. For responses to the questionnaires that *have* been linked to scores on the test, we refer to the following section.

The fact that students say that there are no books in the house may be difficult to accept and indeed this may not be true, technically. Children may simply not be aware of the fact that there are books in the house; they may never see their parents read a book. However, the *factual* presence or the *perception* of the presence of books in the house may be related to the success of students in schools. We will further comment on this below.

Private text books, that is: school books to learn and work from, are available in ca 80% of homes. This may be an important condition for success in education:

students are free to take their books home and study from them when there is a need for this.

Students and teachers seem to disagree as to the extent to which students are made to work in small groups. According to circa one-third of the students this never happens. Project work and practical exercises are part of only minority of lessons. Modern teaching methods, text books and assessment materials emphasize the importance of students working and solving problems in peer groups. For this reason it is worth looking into this issue, given the fact that traditionally Azeri education was frontal and aimed at acquiring and reproducing knowledge, rather than student-centred and competency-based.

The age distribution among teachers is a matter of some concern. As was pointed out, nearly half of the teachers is over 46. However, the youngest age-cohort of 22-25 year-olds makes up only 4% of teachers. If the influx of teachers keeps going down, there will probably be a serious shortage of teachers in 10 years time. It can be imagined that rural schools may suffer most from this tendency. We see similar processes in many other countries in Europe, both in the West and in Central and Eastern Europe.

A similar concern is over the distribution of male and female teachers, with grade 4 teaching being an almost all female occupation. In many countries such a misbalance is to be found, be it not always to the extent found in Azerbaijan. It is obvious that this need not have much influence on student performance, but from a work-relations point of view a more equal representation might be welcome. Azerbaijan teachers are well-qualified and - given their age-distribution - have a considerable work-experience.

Summing up, we recommend activities in the following areas:

- Teacher training, both pre-service and in-service, may have to focus more on group work, project work and on competency-based teaching and learning.
- The teaching profession should be made (more) attractive to younger people, and to women in particular.

11 Relationships of Responses to Questionnaires with Results on Tests

11.1 Summary

In this section we will summarize the results on relationships between background variables and student performance. Some relationships call for more attention than others, because they may seem to signal a significant effect. It may also be the case that such an effect has been expected but was in fact absent. Also, some relationships may be of more of relevance to the present study than others.

Grade 4

Half the sample of fourth-graders is a boy and the other half is a girl. The students of both genders perform equally well on both tests. Students, whose parents have had *higher education*, have higher scores on tests than students whose parents' education stopped at primary or secondary level. The 3% of students at whose homes Azeri is *not* the first language spoken, perform less good on the Mother Tongue test than the others. However, they perform better on the Math test. Students who have more than 110 books at home score better in Math than children with fewer books at home. This effect cannot be seen with Mother Tongue, where one may have expected this. The more books students have read, the better their

scores tend to be on the test for Math. This effect is less pronounced for Mother Tongue.

The availability of such facilities as a room of one's own, private textbooks, computers or calculators does not seem to have much effect on the scores of grade 4 students in Math or Mother Tongue.

The relatively small percentage of students that does not watch TV, performs better in Math than those that do watch TV. The taking part in other leisure activities does not seem to have much effect on scores. Only 8% of grade 4 students has a paid job. These students perform less well on both the Math test and the Mother Tongue test.

The degree to which students like Math and think it is important in life seems to have a positive effect on their scores in Math, but hardly any on their performance in Mother Tongue. Students who find that it is not at all ("absolutely not") difficult to do Math at home, score better than those who find this is much difficult to do at home, but for Mother Tongue this effect has little effect on scores.

Teachers who show an interest in how their students are performing, have students who seem to perform better than students who say their teacher never shows an interest. The same effect is to be seen with teachers who do or do not make efforts to make all students understand a topic. There is some effect to be seen with teachers who do or do not encourage students to solve problems on their own.

Having one's own Math textbooks has some positive effect on scores. The availability of calculators and computers during Math lessons has similar effects. There is less effect with having one's own Mother Tongue textbooks. Little effect can be associated with the availability of magazines and books and of computers and Internet during Mother Tongue lessons.

Grade 9

Half the sample of ninth-graders is a boy and the other half is a girl. The students of both genders perform equally well on both tests. Students whose parents have had *higher education*, have higher scores on tests than students whose parents' education stopped at primary or secondary level. The students at whose homes Azeri is *not* the first language spoken perform less well on the Mother Tongue test than the others. They perform less on the Math test as well, but the effect is less pronounced. Students who have more than 25 books at home score better in Math and Mother Tongue than children with fewer books at home. Reading non-school books has some effect, with the maximum positive effect being reached at 3-5 hours of reading per day.

About one-quarter of grade 9 students has a paid job. These students perform less well on the Math test. For the scores on the Mother Tongue test the effect seems negligible.

Students' jobs in the house for which money is received have some effect on the scores of students.

Taking part in leisure activities such as playing computer games and chatting with friends does not seem to have much effect on scores. Preparing for lessons does have an effect: in the case of Math, students perform best when they spend this much time. In the case of Mother Tongue the effect reaches its maximum when students work 3-5 hours. Doing sport has some effect on the results for Math, with the maximum reached with less than an hour per day. For Mother Tongue these effects are less pronounced: 1-5 hours of sport a day has yielded best results. The degree to which students like Math or Mother Tongue seems to have a positive effect on their scores, for Math stronger than for Mother Tongue. A similar effect is to be seen with the degree to which students agree that these subjects are important in life.

Math teachers who show an interest in how their students are performing, have students who seem to perform better than students who say their teacher never shows an interest. The same effect is to be seen with teachers who do or do not make efforts to make all students understand a topic. There is little effect to be seen with Math teachers who do or do not invite students to explain previous topics studied or who do or do not encourage students to solve problems on their own. Students, whose teachers do or do not encourage them to solve problems on their own, do not show differences in achievement on the test. Other teaching methods as listed in the table above do not show a consistent pattern of effects on the scores of students.

Mother Tongue teachers who show an interest in how their students are performing, have students who seem to perform better than students who say their teacher never shows an interest. There is little effect to be seen with Mother Tongue teachers who do or do not make efforts to make all students understand a topic, or who do or do not invite students to explain previous topics studied. There is some positive effect to be seen with Mother Tongue teachers who encourage students to solve problems on their own. Other teaching methods as indicated in the questionnaire do not show a consistent pattern of effects on the scores of students.

Linking lessons to everyday life seems to have some positive effect on the scores of students.

Whether students have their own Math textbooks or not, has a small positive effect on their scores. The same goes for the Mother Tongue lessons. The availability of calculators and computers during Math lessons has had little or no effect on scores. There is little effect associated with the availability of magazines and books and of computers and Internet during Mother Tongue lessons.

11.2 Discussion and Recommendations

It is important to see that for both grades there seems to be a positive effect on scores where students' parents have had higher education. Thus the positive effects that education can have for the present welfare of the State can be expected to have its effects in a later generation as well. In this context it is a good thing to find that girls and boys take part in education in equal numbers.

The presence of books in the house and certainly reading (non-school) books has some positive effect on scores, more so in grade 9 than in grade 4. It is remarkable that this effect is stronger with Math scores than with Mother Tongue scores, where one would expect such an effect much more. Of course students are taught to read in Language and Literature classes and there are reading activities in class. However, the emphasis in language learning may have been on the formal parts of the language and the formal and historical aspects of Literature. Possibly more emphasis can be put on increasing the number of texts and books that a student must read during his school career, inside and outside school. Teachers may also like to stimulate the enjoyment of reading.

It will be reassuring to for teachers to see that the interest that teachers show in their students, both in grade 4 and in grade 9, seems to have a positive effect on the results of students.

With grade 9 students we have found that working for a paid job inside or outside the home relates with lower scores on Math. It may be worth studying this in more detail. In many economies 15 and 16-year-olds have paid jobs. The possible economic necessity of having such a job, but also the positive sides of having a part-time job may have to be set against the negative effects it can have in the formative years of students, especially when they will not continue their schooling after grade 10 or 12.

Summing up, we recommend activities in the following areas:

- Teaching methods may be developed in which students are encouraged to read and to enjoy reading. Where reading is not promoted at homes, school libraries may help to achieve the aim of encouraging students to read.
- The effect of working for a paid job may have to be studied further.